PDF fitting t-tbar data- experimental considerations

A M Cooper Sarkar

Snowmass Sep 2021

Much of this work is public in
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2018-017/

The importance of the correlation of systematic uncertainties:

- between data points within a spectrum
- between different spectra from a single analysis
- between different spectra from different analyses, different processes

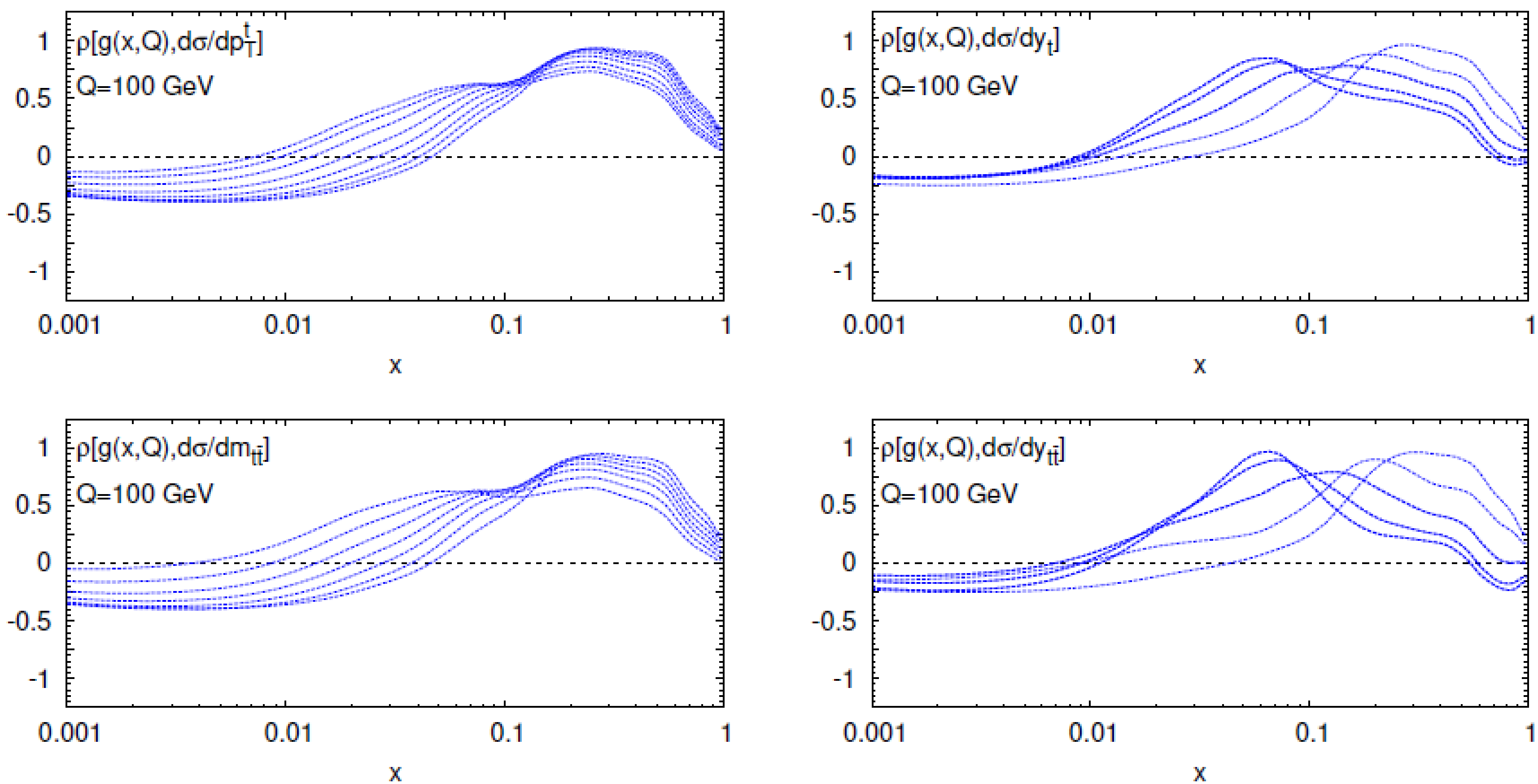
As an example let's discuss fits to

Lepton+jets 8 TeV data from arXIV:1511.04716

https://www.hepdata.net/record/84154

The most constraining top distributions are p_T^t , y_t y_{ttbar}, m_{ttbar} and they mostly constrain the high-x gluon

Here correlation coefficients for each bin of each spectrum with the gluon PDF are plotted as a function of x (from arXIV:1611.08609)



$$\chi^2 = \sum_{ik} \left(D_i - T_i (1 - \sum_i \gamma_{ij} b_j) \right) C^{-1}_{\mathrm{stat},ik}(D_i, D_k) \left(D_k - T_k (1 - \sum_i \gamma_{kj} b_j) \right) + \sum b_j^2$$

How do we actually determine PDFs?
We fit data D, to predictions of NNLO QCD, T, (these predictions rely on the PDFs, which are usually parametrised at an input scale**), taking into account the uncorrelated and correlated uncertainties of the data.**

Uncorrelated is easy, there are statistical and uncorrelated systematics in the Matrix C_{stat} and the statistical component may be bin to bin correlated,
Correlated uncertainties are supplied as fractional, γ, and can be applied as fractions of either data D or theory T, by using nuisance parameters b, which are ideally zero but vary ~± 1 for 1σ variations. These parameters are fitted along with the parameters which describe the PDFs that are input to the predictions.
(This part of the fit is usually done analytically.)

Experimentalists spend YEARS determining the systematic uncertainties of our data. We do the best we can.
But the formalism above assumes systematic uncertainties are well behaved Gaussian errors

They aren’t

First let's consider statistical uncertainties.

The most constraining top distributions are p_T^t , y_t y_{ttbar}, m_{ttbar} and they mostly constrain the high-x gluon

But they can only be fitted simultaneously for maximal information if statistical correlations between (as well as within) the spectra are provided

The statistical correlation matrices within/between the spectra have been evaluated

1	1	0.523003	-0.0865203	-0.230593	-0.123318	-0.0355471	-0.0183453	-0.00742052	0.193366	0.199054	0.165572	0.119437	0.092385	0.191687	0.211977	0.170986	0.141897	0.0801098	0.313266	0.358684	0.1566	0.0379843	-0.00850053	-0.0167743	-0.00412765
2	0.523003	1	0.258974	-0.232733	-0.166392	-0.0568362	-0.0232129	-0.00928392	0.258845	0.259962	0.215854	0.157379	0.11635	0.257006	0.265854	0.221259	0.190785	0.114189	0.367823	0.46559	0.252227	0.0946659	0.0174152	-0.0106161	-0.00640371
3	-0.0865203	0.258974	1	0.211745	-0.16322	-0.0930239	-0.0348096	-0.015387	0.252667	0.256505	0.204611	0.153019	0.112391	0.245568	0.26638	0.217244	0.176239	0.115662	0.206166	0.396708	0.357748	0.205561	0.0849696	0.0238164	-0.00260146
4	-0.230593	-0.232733	0.211745	1	0.189977	-0.128108	-0.057691	-0.0144478	0.174912	0.186568	0.143437	0.121854	0.0838061	0.179544	0.180693	0.151204	0.137084	0.0904561	-0.0151066	0.111899	0.30052	0.306824	0.192282	0.0875043	0.0213644
5	-0.123318	-0.166392	-0.16322	0.189977	1	0.146452	-0.11266	0.028564	0.138193	0.144556	0.113236	0.0888044	0.0533645	0.137564	0.147322	0.115839	0.10241	0.0579249	-0.047342	-0.0350948	0.114799	0.280595	0.292134	0.178125	0.0566404
6	-0.0355471	-0.0568362	-0.0930239	-0.128108	0.146452	1	0.104131	-0.0939535	0.0936114	0.089112	0.0721004	0.0567852	0.0347566	0.0933875	0.0967962	0.0795374	0.0556884	0.0343439	-0.0254326	-0.0447006	-0.0102559	0.101504	0.262709	0.226378	0.0966235
7	-0.0183453	-0.0232129	-0.0348096	-0.057691	-0.112662	0.104131	1	0.0584419	0.0583097	0.0677387	0.0534449	0.028248	0.0197735	0.0582822	0.065875	0.0532866	0.0378831	0.0245572	-0.0164502	-0.0264189	-0.0266078	0.00383807	0.112064	0.272357	0.163028
8	-0.00742052	-0.00928392	-0.015387	-0.0144478	-0.028564	-0.0939535	0.0584419	1	0.0351814	0.044628	0.0329678	0.0154428	0.0038405	0.0365307	0.0394551	0.0347108	0.0206278	0.00892468	0.000869021	-0.00799165	-0.0166962	-0.0131964	0.0139482	0.158587	0.205627
1	0.193366	0.258845	0.252667	0.174912	0.138193	0.0936114	0.0583097	0.0351814	1	-0.0196275	-0.209871	-0.138556	-0.058594	0.489861	0.299719	0.0681119	-0.046792	-0.0606758	0.171685	0.262747	0.237413	0.19099	0.144966	0.0966219	0.0493263
2	0.199054	0.259962	0.256505	0.186568	0.144556	0.089112	0.0677387	0.044628	-0.0196275	1	0.0361891	-0.139761	-0.090291	0.247887	0.3457	0.236424	0.0775936	-0.0404607	0.169373	0.273609	0.248759	0.201697	0.147774	0.0990592	0.0480445
3	0.165572	0.215854	0.204611	0.143437	0.113236	0.0721004	0.0534449	0.0329678	-0.209871	0.0361891	1	0.055808	-0.116518	0.0399956	0.170229	0.273071	0.257601	0.0579035	0.132723	0.21484	0.193211	0.156733	0.12442	0.102127	0.0485058
4	0.119437	0.157379	0.153019	0.121854	0.0888044	0.0567852	0.028248	0.0154428	-0.138556	-0.139761	0.055808	1	0.0230257	-0.0363669	0.0317	0.158325	0.296765	0.255948	0.0878424	0.130365	0.141035	0.130769	0.120789	0.0846287	0.0346981
5	0.092385	0.11635	0.112391	0.0838061	0.0533645	0.0347566	0.0197735	0.0038405	-0.058594	-0.090291	-0.116518	0.0230257	1	-0.0239123	-0.0160335	0.0398061	0.185696	0.379725	0.0457679	0.0820693	0.0974089	0.10323	0.0949771	0.0793067	0.0358952
1	0.191687	0.257006	0.245568	0.179544	0.137564	0.0933875	0.0582822	0.0365307	0.489861	0.247887	0.0399956	-0.0363669	-0.0239123	1	-0.0274769	-0.263436	-0.0870922	0.00998395	0.134049	0.225154	0.221049	0.203422	0.183829	0.144336	0.0816254
2	0.211977	0.265854	0.26638	0.180693	0.147322	0.0967962	0.065875	0.0394551	0.299719	0.3457	0.170229	0.0317	-0.0160335	-0.0274769	1	0.17176	-0.195766	-0.0497008	0.158414	0.257381	0.246893	0.20935	0.176059	0.136116	0.0727846
3	0.170986	0.221259	0.217244	0.151204	0.115839	0.0795374	0.0532866	0.0347108	0.0681119	0.236424	0.273071	0.158325	0.0398061	-0.263436	0.17176	1	0.111911	-0.180419	0.132604	0.217563	0.209858	0.173215	0.13443	0.100188	0.0396299
4	0.141897	0.190785	0.176239	0.137084	0.10241	0.0556884	0.0378831	0.0206278	-0.046792	0.0775936	0.257601	0.296765	0.185696	-0.0870922	-0.195766	0.111911	1	-0.0338399	0.129643	0.204679	0.185544	0.142059	0.0939908	0.0489947	0.00816948
5	0.0801098	0.114189	0.115662	0.0904561	0.0579249	0.0343439	0.0245572	0.00892468	-0.0606758	-0.0404607	0.0579035	0.255948	0.379725	0.00998395	-0.0497008	-0.180419	-0.0338399	1	0.117539	0.140678	0.099773	0.069673	0.0272868	0.00203619	-0.00630476
1	0.313266	0.367823	0.206166	-0.0151066	-0.047342	-0.0254326	-0.0164502	0.000869021	0.171685	0.169373	0.132723	0.0878424	0.0457679	0.134049	0.158414	0.132604	0.129643	0.117539	1	0.52057	-0.254831	-0.249306	-0.10418	-0.0214541	0.000815721
2	0.358684	0.46559	0.396708	0.111899	-0.0350948	-0.0447006	-0.0264189	-0.00799165	0.262747	0.273609	0.21484	0.130365	0.0820693	0.225154	0.257381	0.217563	0.204679	0.140678	0.52057	1	0.298679	-0.250712	-0.221548	-0.0669809	-0.010135
3	0.1566	0.252227	0.357748	0.30052	0.114799	-0.0102559	-0.0266078	-0.0166962	0.237413	0.248759	0.193211	0.141035	0.0974089	0.221049	0.246893	0.209858	0.185544	0.099773	-0.254831	0.298679	1	0.409987	-0.163016	-0.142808	-0.0315226
4	0.0379843	0.0946659	0.205561	0.306824	0.280595	0.101504	0.00383807	-0.0131964	0.19099	0.201697	0.156733	0.130769	0.10323	0.203422	0.20935	0.173215	0.142059	0.069673	-0.249306	-0.250712	0.409987	1	0.334172	-0.169248	-0.0742702
5	-0.00850053	0.0174152	0.0849696	0.192282	0.292134	0.262709	0.112064	0.0139482	0.144966	0.147774	0.12442	0.120789	0.0949771	0.183829	0.176059	0.13443	0.0939908	0.0272868	-0.10418	-0.221548	-0.163016	0.334172	1	0.184151	-0.122152
6	-0.0167743	-0.0106161	0.0238164	0.0875043	0.178125	0.226378	0.272357	0.158587	0.0966219	0.0990592	0.102127	0.0846287	0.0793067	0.144336	0.136116	0.100188	0.0489947	0.00203619	-0.0214541	-0.0669809	-0.142808	-0.169248	0.184151	1	0.145869
7	-0.00412765	-0.00640371	-0.00260146	0.0213644	0.0566404	0.0966235	0.163028	0.205627	0.0493263	0.0480445	0.0485058	0.0346981	0.0358952	0.0816254	0.0727846	0.0396299	0.00816948	-0.00630476	0.000815721	-0.010135	-0.0315226	-0.0742702	-0.122152	0.145869	1
	1	2	3	4	5	6	7	8	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	6	7

Table 1: Statistical correlation matrix between the absolute differential cross-sections. All variables are included to show the correlations between different bins of different variables. From left to right and top to bottom the rows and columns are labeled by bin number for each variable and the variables are ordered: p_T^t, $|y_t|$, $|y_{t\bar{t}}|$, and $m_{t\bar{t}}$.

This information is added to the HEPDATA entry for the lepton+jets spectra

Tables 167,168,169,170,172,173,174,176,177,179

https://www.hepdata.net/record/84154

Tables 29,31,27,23 for the distributions themselves

Now let's consider systematic uncertainties.

MANY of these are correlated bin to bin both within and between spectra.
In particular, some systematic uncertainties are what are referred to as
'2-point systematics'
This means they are determined by running one Monte-Carlo data simulator, say PYTHIA, and another, say HERWIG, and taking the difference as the systematic uncertainty. This is a reasonable estimate, it is not a Gaussian error.
Unfortunately, such uncertainties are often the largest systematics--- more than a few percent.

The formalism also assumes correlated systematic uncertainties are 100% correlated point to point throughout the data set to which they apply.
100% may not be realistic.
AND it has become common practice to assign more and more systematics.
AT HERA we had 169 for ~1200 data points
AT LHC we often have >~300 for <~300 data points (in some cases MUCH less data)
So we had better be treating them right.

AN example:

ATLAS data on t-tbar differential distribution

As an example let's discuss fits to

t-tbar differential distributions in lepton+jets channel at 8 TeV data from

arXIV:1511.04716

https://www.hepdata.net/record/84154

Top data exists as normalised and absolute spectra .
Absolute also carries information on the total t-tbar cross-sections which is useful to constrain PDF fits.
We will consider absolute spectra but the considerations are similar for normalised spectra - although the particular systematic uncertainties that are important may differ

- For the specific fits used in these examples, the top data are used in addition to the HERA I+II combined data, and the ATLAS W,Z 7 TeV data
- The top data and W,Z data are complementary – top affects the gluon, whereas W,Z affects the quarks.
- Conclusions on top are similar if W,Z is removed

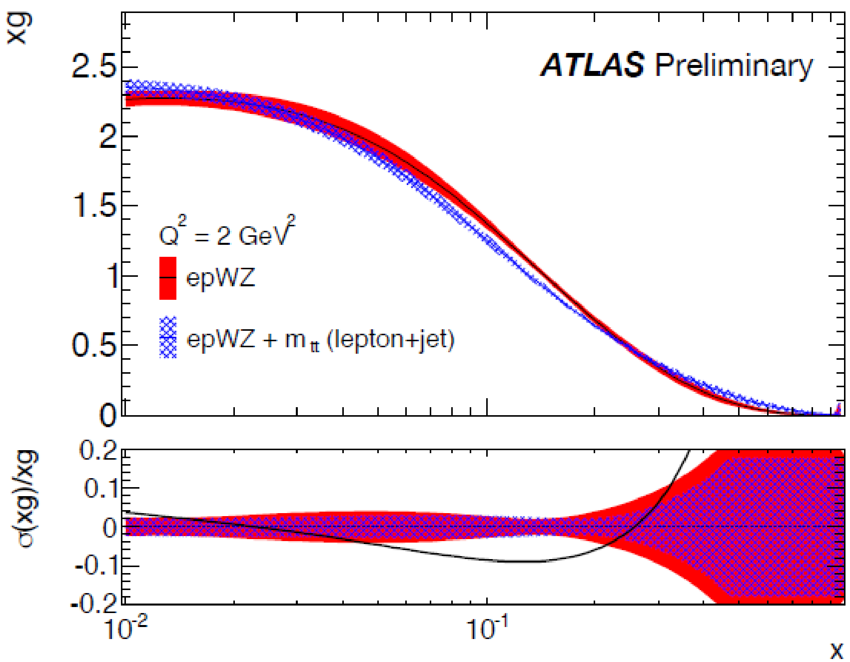
There is no tension between the top data and the other data sets in the fit
Note global fits have many more data sets, which could be in tension with the these data, notably jet data.

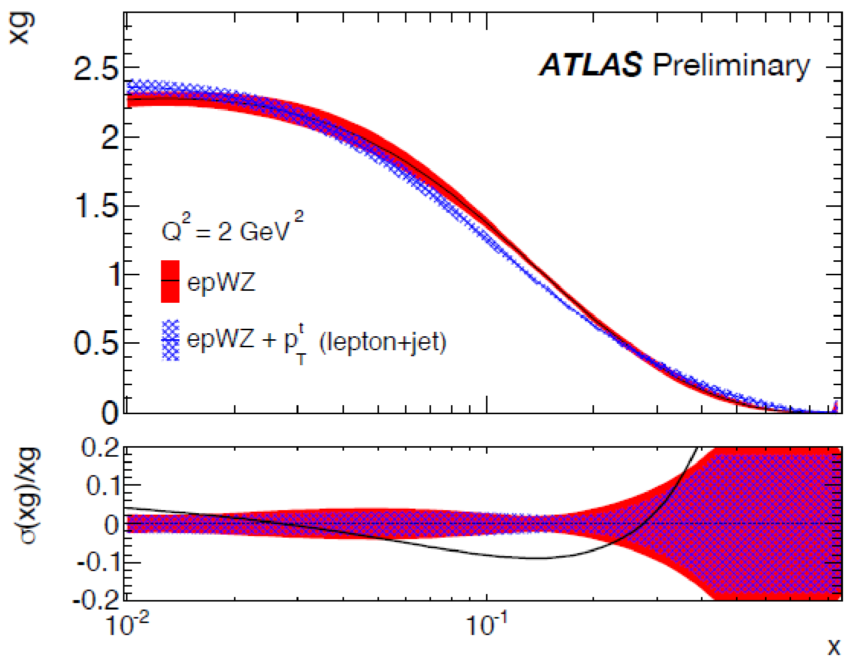
First consider one spectrum at a time

		lepton+jets spectrum			
		$m_{t\bar{t}}$	p_T^t	$y_{t\bar{t}}$	y_t
Total χ^2/NDF		1238.4 / 1062	1239.4 / 1063	1257.5 / 1060	1246.5 / 1060
Partial χ^2/NDP	HERA	1153 / 1016	1151 / 1016	1149 / 1016	1146 / 1016
Partial χ^2/NDP	ATLAS $W, Z/\gamma^*$	82.0 / 55	82.1 / 55	86.4 / 55	85.0 / 55
Partial χ^2/NDP	ATLAS $t\bar{t}$	3.4 / 7	7.9 / 8	19.7 / 5	18.3 / 5

χ2 for p_T^t and m_{ttbar} are good

The χ2 for the HERA and ATLAS W,Z are similar to when they are fitted without top—there is no tension





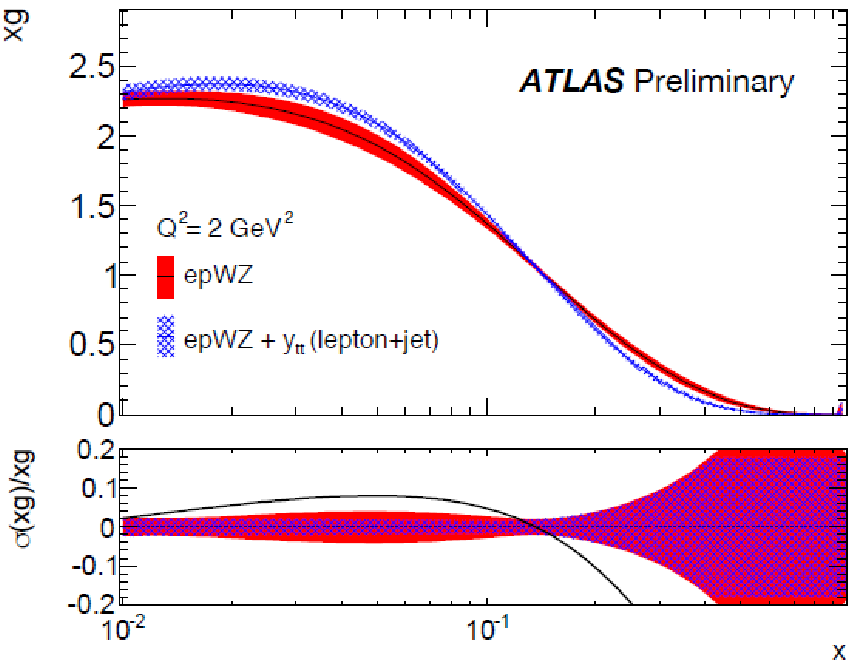
Both p_T^t and m_{ttbar} spectra harden the gluon in comparison to just ATLAS epWZ (HERA +ATLAS WZ2011)

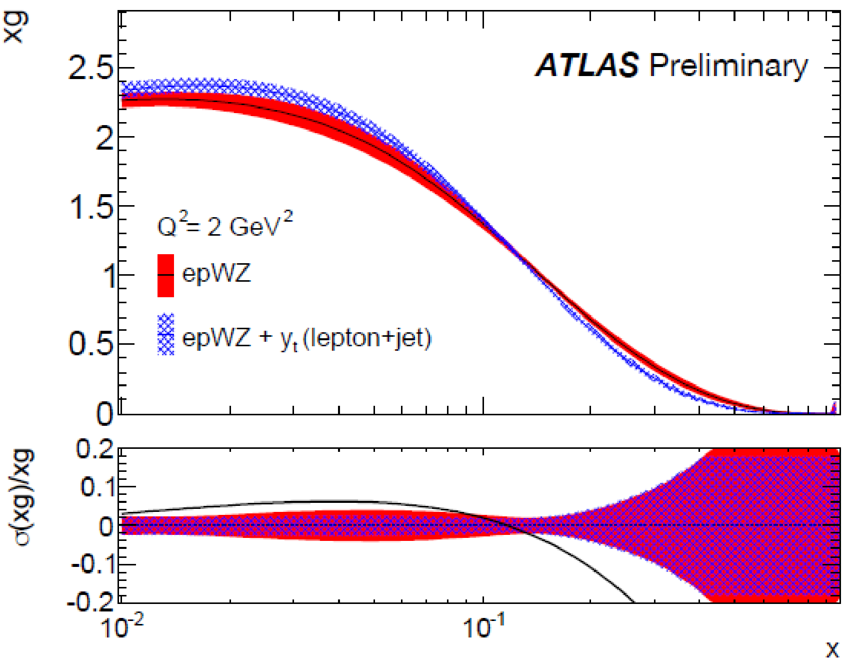
Now consider one spectrum at a time

		lepton+jets spectrum			
		m_{tt}	p_T^t	y_{tt}	y_t
Total χ^2/NDF		1238.4 / 1062	1239.4 / 1063	1257.5 / 1060	1246.5 / 1060
Partial χ^2/NDP	HERA	1153 / 1016	1151 / 1016	1149 / 1016	1146 / 1016
Partial χ^2/NDP	ATLAS $W, Z/\gamma^*$	82.0 / 55	82.1 / 55	86.4 / 55	85.0 / 55
Partial χ^2/NDP	ATLAS $t\bar{t}$	3.4 / 7	7.9 / 8	19.7 / 5	18.3 / 5

χ2 for y_t and y_{ttbar} are not good.

The χ2 for the HERA and ATLAS W,Z are similar to when they are fitted without top—there is no tension.





Both y_t and y_{ttbar} spectra soften the gluon

NOW try fitting 2 spectra at a time: (p_T^t and y_t) and (p_T^t and m_{ttbar})
-------------------------look at the χ2 for these fits

		lepton+jets spectra			
		p_T^t and y_t with statistical correlations	p_T^t and y_t without statistical correlations	p_T^t and m_{tt} with statistical correlations	p_T^t and m_{tt} without statistical correlations
Total χ^2/NDF		1264 / 1068	1260 / 1068	1290 / 1070	1287 / 1070
Partial χ^2/NDP	HERA	1148 / 1016	1147 / 1016	1162 / 1016	1162 / 1016
Partial χ^2/NDP	ATLAS $W,Z/\gamma^*$	82.7 / 55	83.5 / 55	83.2 / 55	83.1 / 55
Partial χ^2/NDP	ATLAS $t\bar{t}$	33 / 13	30 / 13	45 / 15	42 / 15

This Table shows fits to (p_T^t and y_t) and (p_T^t and mtt) simultaneously.
In all cases the correlated systematics between the spectra are included.
The correlated statistical uncertainties are used by default but are also switched off to assess their impact. This makes it clear that the statistical correlations are NOT the source of the bad χ2

None of these top χ2 is satisfactory BUT the p_T^t + y_t χ2 is only a bit larger than the added sum of the p_T^t and y_t separate fit χ2 = 26.2, so the main problem here is the poor fit to y_t
whereas the p_T^t +m_{ttbar} χ2 is much larger than the sum of the p_T^t and m_{ttbar} separate χ2 = 11.3-
This is surprising since the fits to the individual spectra are good

Since the source of the poor χ2 is NOT the statistical correlations we look at the systematic correlations. Should they ALL be correlated between the spectra?

Three particularly LARGE systematic uncertainties are the sys isr/fsr (~8%) and the sys-ps_model (~5%) and the hard scattering model (~4%). These are '2-point systematics'. Let's look at the fitted values of the nuisance parameters, b, for these 3 systematic uncertainties, when they are fitted separately

Systematic uncertainty source	lepton+jets spectrum			
	p_T^t	y_t	y_{tt}	m_{tt}
Hard scattering model	+0.74± 0.31	+0.48± 0.22	+0.92± 0.37	-0.43± 0.20
Parton shower model	-1.32± 0.43	-0.79± 0.26	-0.51± 0.17	+0.39±0.13
ISR/FSR model	-0.47± 0.18	-0.87±0.30	-1.27± 0.38	+0.33± 0.10

$$\chi^2 = \sum_{ik} \left(D_i - T_i(1 - \sum_j \gamma_{ij} b_j) \right) C^{-1}_{\mathrm{stat},ik}(D_i, D_k) \left(D_k - T_k(1 - \sum_j \gamma_{kj} b_j) \right) + \sum_j b_j^2$$

The treatment of correlated systematics as nuisance parameters means that they can introduce correlated shifts in the predictions. Examining the shifts due to these 3 sources shows that the m_{ttbar} spectrum induces an opposite shift to the other three spectra, when the spectra are fitted separately. When fitting together the shifts are forced to be the same **---if 100% correlation is assumed between the spectra**. E.g. the common nuisance parameter for the Parton Shower uncertainty when fitting p_T^t and m_{ttbar} together is -0.32 ± 0.10, which suits neither spectrum.

Let's investigate decorrelating these sources of systematic uncertainty between the spectra, while preserving bin-to-bin correlations within the spectra.
First decorrelate all 3 sources simultaneously and then decorrelate one at a time.
This shows us that it is the decorrelation of the parton shower systematic which is the most significant (with the isr/fsr uncertainty a close second)

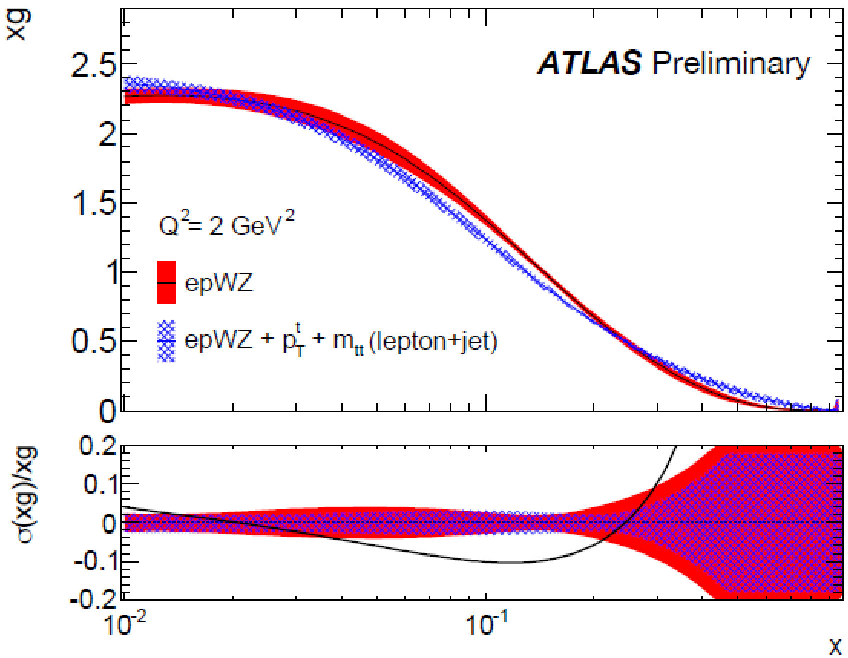
		lepton+jets spectra		
		p_T^t and y_t decorrelate 2-point uncertainties	p_T^t and m_{tt} decorrelate 2-point uncertainties	p_T^t and m_{tt} decorrelate parton-shower model uncertainty
Total χ^2/NDF		1259 / 1068	1247 / 1070	1248 / 1070
Partial χ^2/NDP	HERA	1147 / 1016	1154 / 1016	1153 / 1016
Partial χ^2/NDP	ATLAS $W, Z/\gamma^*$	83.9 / 55	81.9 / 55	81.6 / 55
Partial χ^2/NDP	ATLAS $t\bar{t}$	27.8 / 13	11.5 / 15	14.1 / 15

The effect of decorrelation is marginal for the p_T^t and y_t spectra, as expected since the shifts induced by these spectra are similar when they are fitted separately. The resultant χ2 is closer to the sum of the χ2 of the separate fits (26.2) but is not changed much
The effect of decorrelation is dramatic for the p_T^t and m_{ttbar} spectra, now that the shifts are allowed to be different. (The separate nuisance parameters are -0.47 for pt and +0.10 for mtt). The resultant χ2 is close to the sum of the χ2 of the separate fits (11.3)

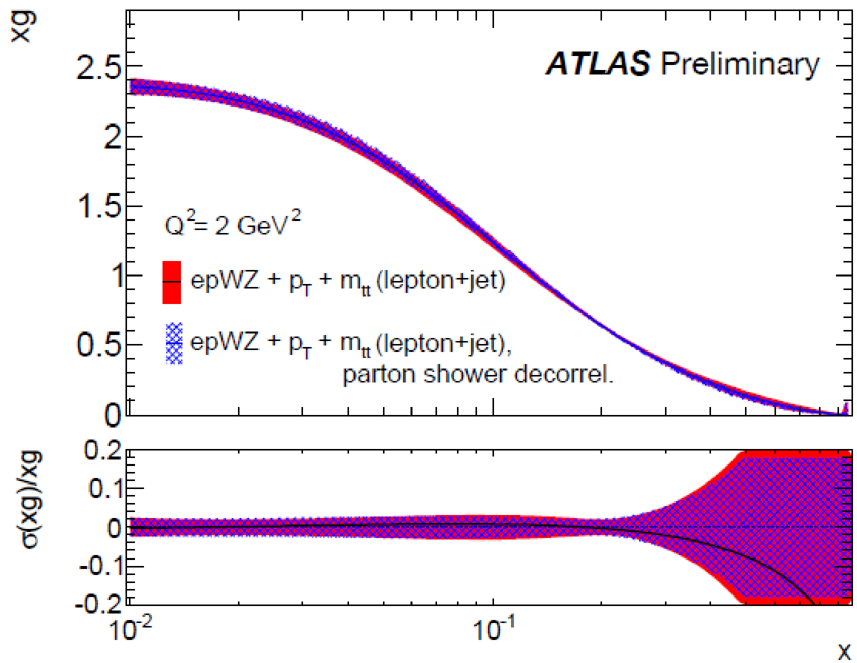
The resultant shape of the gluon barely changes when these systematics are decorrelated- the main effect is the improvement in χ2

The resultant shape of the gluon barely changes when these systematics are decorrelated- the main effect is the improvement in χ2

All uncertainties fully correlated



Compare parton shower uncertainty correlated/decorrelated



100% correlation has a marginally stronger pull on the gluon and a marginally smaller uncertainty.

We chose to decorrelate the parton shower systematic uncertainty between the spectra. This choice has now also been made by CT and similar choices are made by MSHT. **But you should only do what the experimentalists will support.**

The freedom to do this is WHY we want the information on systematic uncertainty separated into its many sources with preserved bin to bin signs….

What is the most useful way to present the information on correlated uncertainties?
The form of the χ2 used treats correlated systematics in terms of nuisance parameters b_j

$$\chi^2 = \sum_{ik} \left(D_i - T_i(1 - \sum_j \gamma_{ij} b_j) \right) C^{-1}_{\mathrm{stat},ik}(D_i, D_k) \left(D_k - T_k(1 - \sum_j \gamma_{kj} b_j) \right) \;\; + \sum_j b_j^2$$

The correlated systematic uncertainties and their correlations between different bins (the γ^i_j terms, typically obtained from $\frac{\mu^i_{j,\mathrm{up}} - \mu^i_{j,\mathrm{down}}}{2\mu^i_{\mathrm{nom}}}$ preserving the sign of the uncertainty in each bin)

Example of good practice

$\frac{1}{\sigma}\frac{d\sigma}{dp^t_{\mathrm{T}}}$ Uncertainties [%] / Bins [GeV]	0–50	50–100	100–150	150–200	200–250	250–350	350–800
b-quark jets (JES)	$^{-1.81}_{+1.52}$	$^{-0.41}_{+0.55}$	$^{+1.12}_{-0.97}$	$^{+0.89}_{-1.10}$	$^{+0.46}_{-0.52}$	$^{+0.07}_{-0.23}$	$^{-0.01}_{-0.11}$
Close-by jets (JES)	$^{-1.56}_{+1.45}$	$^{-0.45}_{+0.61}$	$^{+0.40}_{-0.77}$	$^{+1.39}_{-1.01}$	$^{+1.13}_{-1.02}$	$^{+1.47}_{-1.39}$	$^{+1.80}_{-1.65}$
Effective detector NP set 1 (JES)	$^{-1.51}_{+1.06}$	$^{-0.47}_{+0.54}$	$^{+0.54}_{-0.40}$	$^{+1.16}_{-1.20}$	$^{+1.20}_{-1.04}$	$^{+1.17}_{-0.97}$	$^{+1.49}_{-0.51}$
Effective detector NP set 2 (JES)	$^{-0.21}_{+0.09}$	$^{+0.00}_{+0.05}$	$^{-0.06}_{-0.08}$	$^{+0.24}_{-0.06}$	$^{+0.23}_{+0.05}$	$^{+0.12}_{-0.18}$	$^{+0.28}_{-0.26}$
Effective mixed NP set 1 (JES)	$^{-0.03}_{+0.04}$	$^{-0.02}_{+0.11}$	$^{-0.07}_{-0.12}$	$^{+0.11}_{-0.09}$	$^{+0.07}_{-0.10}$	$^{+0.23}_{+0.04}$	$^{+0.79}_{+0.34}$
Effective mixed NP set 2 (JES)	$^{-0.03}_{+0.03}$	$^{-0.03}_{+0.12}$	$^{+0.04}_{-0.25}$	$^{-0.01}_{+0.07}$	$^{+0.04}_{+0.00}$	$^{+0.08}_{+0.18}$	$^{+0.24}_{+0.85}$
Effective model NP set 1 (JES)	$^{+0.28}_{+0.05}$	$^{+0.02}_{+0.04}$	$^{-0.28}_{-0.22}$	$^{+0.30}_{-0.24}$	$^{+0.19}_{+0.48}$	$^{-0.64}_{+0.85}$	$^{-1.40}_{+1.51}$
Effective model NP set 2 (JES)	$^{-0.53}_{+0.59}$	$^{-0.10}_{+0.17}$	$^{-0.11}_{-0.22}$	$^{+0.62}_{-0.38}$	$^{+0.63}_{-0.51}$	$^{+0.76}_{-0.48}$	$^{+1.34}_{-0.48}$
Effective model NP set 3 (JES)	$^{+0.66}_{-0.80}$	$^{+0.23}_{-0.14}$	$^{-0.34}_{+0.05}$	$^{-0.44}_{+0.69}$	$^{-0.46}_{+0.84}$	$^{-0.42}_{+0.77}$	$^{-0.28}_{+0.88}$
Effective model NP set 4 (JES)	$^{-0.14}_{+0.16}$	$^{+0.03}_{+0.02}$	$^{-0.06}_{-0.12}$	$^{+0.08}_{+0.03}$	$^{+0.21}_{-0.06}$	$^{+0.10}_{-0.05}$	$^{+0.23}_{+0.10}$
Effective statistical NP set 1 (JES)	$^{+0.52}_{-0.65}$	$^{+0.26}_{-0.07}$	$^{-0.39}_{+0.02}$	$^{-0.18}_{+0.39}$	$^{-0.20}_{+0.61}$	$^{-0.82}_{+0.92}$	$^{-1.16}_{+1.53}$
Effective statistical NP set 2 (JES)	$^{-0.20}_{+0.21}$	$^{+0.04}_{+0.06}$	$^{-0.05}_{-0.14}$	$^{+0.15}_{-0.07}$	$^{+0.10}_{-0.12}$	$^{+0.19}_{-0.14}$	$^{+0.49}_{+0.04}$
Effective statistical NP set 3 (JES)	$^{-0.44}_{+0.43}$	$^{-0.14}_{+0.18}$	$^{+0.16}_{-0.30}$	$^{+0.33}_{-0.20}$	$^{+0.39}_{-0.28}$	$^{+0.32}_{-0.30}$	$^{+0.46}_{-0.02}$

Information in this form is much more useful than a covariance matrix when there is an issue with systematics—and there often is

The uncertainties don't HAVE to be asymmetric BUT they do HAVE to be SIGNED to be useful

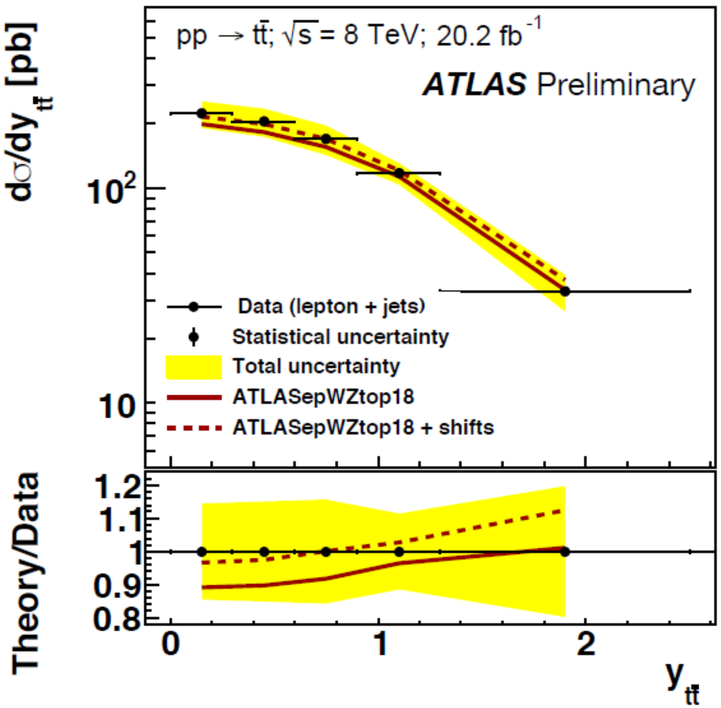
Correlations BETWEEN the different t-tbar spectra; ptavt, mtt etc, are easily dealt with because the systematic uncertainties carry the same names– However, please pay attention to consistency of sign

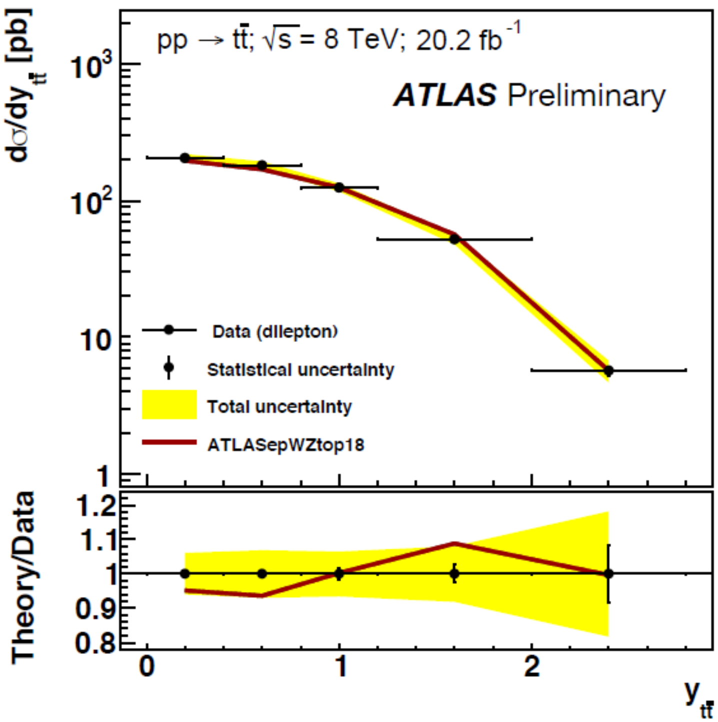
Those paying close attention will have noticed that the rapidity distributions are still not well fitted.

Compare y_{ttbar} from lepton+jets and dilepton channels to the predictions of this fit

There is a trend of the y_{ttbar} lepton+jets data that is hard to fit despite comparable level of total uncertainties

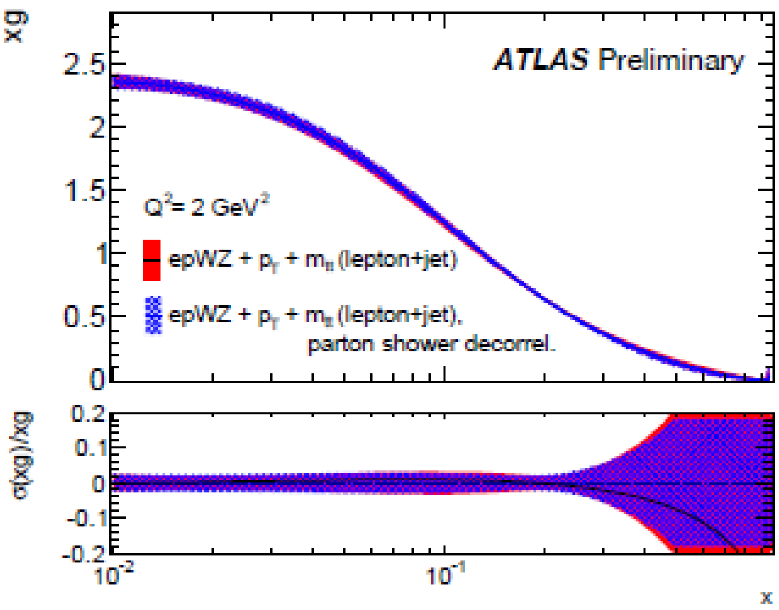
For the dilepton channel statistical (uncorrelated) uncertainties are a larger contribution to the total --- correlated systematic uncertainties matter

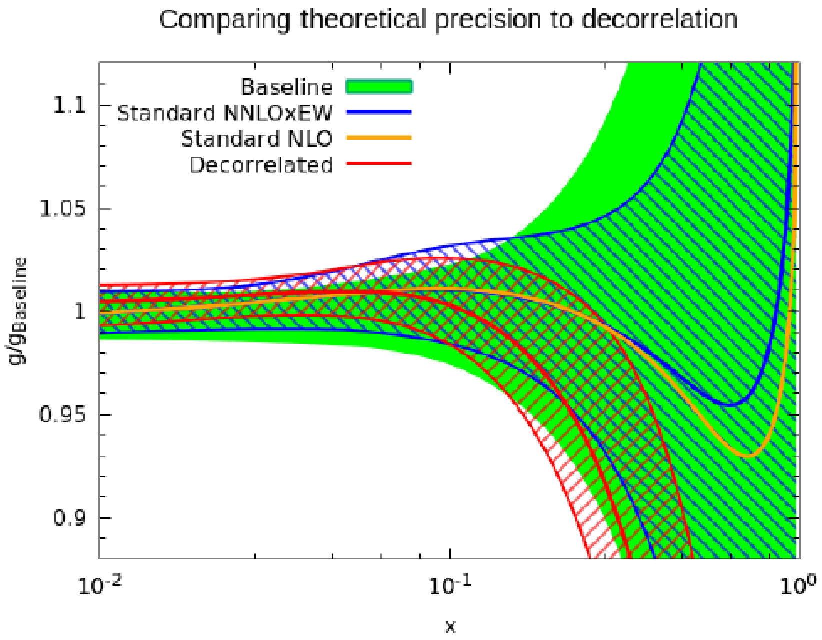




This suggests that one MIGHT need decorrelation within the ytt spectrum ie differing b parameters as a function of ytt

This was done by the MMHT group in arXiv:1909.10541.
In the simple ATLAS study decorrelating ONLY parton shower between p^t_T and m_{tt}.
the effect of decorrelation is not very significant.
But for the arxiv:1909.10541 study decorrelating parton shower between all 4 spectra and using decorrelation within the rapidity spectra we see that the effect can be larger than the NLO to NNLO difference.





Parting warnings

There is a further problem of relating the source of systematic uncertainty between two different types of data. Could we have a naming convention?

e.g. Using ATLAS W+jets from 1711.03296 and Z+jets from 1907.06728
For Z+jets we have uncertainties called 'ATL_unfold_Data' and 'ATL_unfold_MC'
For W+jets these are called 'UnfoldReweight' and 'UnfoldOtherGen'
This is not so hard --one can talk to the authors, one can read the papers carefully BUT as time goes on memory is lost.

WORSE STILL, the more data we add the more correlations we need to consider
For example consider V +jets final states and t-tbar in the lepton+jets channels
There could be correlations in the jet systematics between these channels and of both of these channels to the inclusive jets.
And the JET systematics are the largest

These inter-data-set correlations are not taken into account in any PDF fits.
We experimentalist could try to be more helpful in providing inter-data-set correlations in a readily understandable form.

If we want 1% accuracy on PDFs this matters!

Recently the statistical correlation matrices between the spectra have been evaluated

1	1	0.523003	-0.0865203	-0.230593	-0.123318	-0.0355471	-0.0183463	-0.00742052	0.193366	0.199054	0.165572	0.119437	0.092385	0.191687	0.211977	0.170986	0.141897	0.0801098	0.313266	0.358684	0.1566	0.0379843	-0.00850053	-0.0167743	-0.00412765
2	0.523003	1	0.258974	-0.232733	-0.166202	-0.0568362	-0.0232129	-0.00928292	0.258845	0.259962	0.215854	0.157379	0.11635	0.257006	0.263854	0.221259	0.190785	0.114189	0.367823	0.46559	0.252227	0.0946659	0.0174152	-0.0106161	-0.00640371
3	-0.0865203	0.258974	1	0.211745	-0.16322	-0.0930239	-0.0348096	-0.015387	0.252667	0.256505	0.204611	0.153019	0.112391	0.245568	0.26638	0.217244	0.176239	0.115662	0.206166	0.296708	0.357748	0.205561	0.0849696	0.0238164	-0.00260146
4	-0.230593	-0.232733	0.211745	1	0.189977	-0.128108	-0.057691	-0.0144478	0.174912	0.186568	0.143437	0.121854	0.0838061	0.179544	0.180693	0.151204	0.137084	0.0904961	-0.0151066	0.111899	0.30052	0.306824	0.192282	0.0875043	0.0213644
5	-0.123318	-0.166202	-0.16322	0.189977	1	0.146452	-0.11266	0.028564	0.138193	0.144556	0.113236	0.0888044	0.0533645	0.137564	0.147322	0.115839	0.10241	0.0579249	-0.047342	-0.0350948	0.114799	0.280595	0.292134	0.178125	0.0566404
6	-0.0355471	-0.0568362	-0.0930239	-0.128108	0.146452	1	0.104131	-0.0939535	0.0936114	0.089112	0.0721004	0.0567852	0.0347566	0.0933875	0.0957962	0.0795374	0.0556884	0.0343439	-0.0254326	-0.0447006	-0.0102559	0.101504	0.262709	0.226378	0.0986235
7	-0.0183463	-0.0232129	-0.0348096	-0.057691	-0.112662	0.104131	1	0.0584419	0.0583097	0.0677387	0.0534449	0.028248	0.0197735	0.0582822	0.065875	0.0532866	0.0378831	0.0245572	-0.0164502	-0.0264189	-0.0266078	0.00383807	0.112064	0.272357	0.163028
8	-0.00742052	-0.00928292	-0.015387	-0.0144478	-0.028564	-0.0939535	0.0584419	1	0.0351814	0.044628	0.0329678	0.0154428	0.0038406	0.0365307	0.0394551	0.0347108	0.0206278	0.00892468	0.000869021	-0.00799165	-0.0166962	-0.0131954	0.0139482	0.158587	0.205627
1	0.193366	0.258845	0.252667	0.174912	0.138193	0.0936114	0.0583097	0.0351814	1	-0.0196275	-0.209871	-0.138556	-0.058594	0.489861	0.299719	0.0681119	-0.046792	-0.0606758	0.171685	0.262747	0.237413	0.19099	0.144966	0.0966219	0.0493263
2	0.199054	0.259962	0.256505	0.186568	0.144556	0.089112	0.0677387	0.044628	-0.0196275	1	0.0361891	-0.139761	-0.090291	0.247887	0.3457	0.236424	0.0775306	-0.0404607	0.169373	0.273609	0.248759	0.201697	0.147774	0.0990592	0.0480445
3	0.165572	0.215854	0.204611	0.143437	0.113236	0.0721004	0.0534449	0.0329678	-0.209871	0.0361891	1	0.055808	-0.116518	0.0309966	0.170229	0.273071	0.257601	0.0579035	0.132723	0.21484	0.193211	0.156733	0.12442	0.102127	0.0485058
4	0.119437	0.157379	0.153019	0.121854	0.0888044	0.0567852	0.028248	0.0154428	-0.138556	-0.139761	0.055808	1	0.0230257	-0.0363669	0.0317	0.158325	0.296765	0.255948	0.0878424	0.130365	0.141035	0.130769	0.120789	0.0846287	0.0346981
5	0.092385	0.11635	0.112391	0.0838061	0.0533645	0.0347566	0.0197735	0.0038406	-0.058594	-0.090291	-0.116518	0.0230257	1	-0.0239123	-0.0160335	0.0398061	0.185696	0.379725	0.0457679	0.0820693	0.0974089	0.10323	0.0949771	0.0793067	0.0358952
1	0.191687	0.257006	0.245568	0.179544	0.137564	0.0933875	0.0582822	0.0365307	0.489861	0.247887	0.0309966	-0.0363669	-0.0239123	1	-0.0274769	-0.263436	-0.0870922	0.00998305	0.134049	0.225154	0.221049	0.203422	0.183829	0.144336	0.0816254
2	0.211977	0.263854	0.26638	0.180693	0.147322	0.0957962	0.065875	0.0394551	0.299719	0.3457	0.170229	0.0317	-0.0160335	-0.0274769	1	0.17176	-0.195766	-0.0497008	0.158414	0.257381	0.246893	0.20935	0.176059	0.136116	0.0727846
3	0.170986	0.221259	0.217244	0.151204	0.115839	0.0795374	0.0532866	0.0347108	0.0681119	0.236424	0.273071	0.158325	0.0398061	-0.263436	0.17176	1	0.111911	-0.180419	0.132604	0.217563	0.209658	0.173215	0.13443	0.100188	0.0396299
4	0.141897	0.190785	0.176239	0.137084	0.10241	0.0556884	0.0378831	0.0206278	-0.046792	0.0775306	0.257601	0.296765	0.185696	-0.0870922	-0.195766	0.111911	1	-0.0338399	0.128643	0.204679	0.185544	0.142059	0.0939908	0.0489947	0.00816948
5	0.0801098	0.114189	0.115662	0.0904961	0.0579249	0.0343439	0.0245572	0.00892468	-0.0606758	-0.0404607	0.0579035	0.255948	0.379725	0.00998305	-0.0497008	-0.180419	-0.0338399	1	0.117539	0.146678	0.099773	0.069673	0.0272868	0.00203619	-0.00530476
1	0.313266	0.367823	0.206166	-0.0151066	-0.047342	-0.0254326	-0.0164502	0.000869021	0.171685	0.169373	0.132723	0.0878424	0.0457679	0.134049	0.158414	0.132604	0.128643	0.117539	1	0.52067	-0.254831	-0.249305	-0.10418	-0.0214541	0.000815721
2	0.358684	0.46559	0.296708	0.111899	-0.0350948	-0.0447006	-0.0264189	-0.00799165	0.262747	0.273609	0.21484	0.130365	0.0820693	0.225154	0.257381	0.217563	0.204679	0.146678	0.52067	1	0.298679	-0.250712	-0.221548	-0.0669809	-0.010135
3	0.1566	0.252227	0.357748	0.30052	0.114799	-0.0102559	-0.0266078	-0.0166962	0.237413	0.248759	0.193211	0.141035	0.0974089	0.221049	0.246893	0.209658	0.185544	0.099773	-0.254831	0.298679	1	0.409987	-0.163016	-0.142808	-0.0315226
4	0.0379843	0.0946659	0.205561	0.306824	0.280595	0.101504	0.00383807	-0.0131954	0.19099	0.201697	0.156733	0.130769	0.10323	0.203422	0.20935	0.173215	0.142059	0.069673	-0.249305	-0.250712	0.409987	1	0.334172	-0.169248	-0.0742702
5	-0.00850053	0.0174152	0.0849696	0.192282	0.292134	0.262709	0.112064	0.0139482	0.144966	0.147774	0.12442	0.120789	0.0949771	0.183829	0.176059	0.13443	0.0939908	0.0272868	-0.10418	-0.221548	-0.163016	0.334172	1	0.184151	-0.122152
6	-0.0167743	-0.0106161	0.0238164	0.0875043	0.178125	0.226378	0.272357	0.158587	0.0966219	0.0990592	0.102127	0.0846287	0.0793067	0.144336	0.136116	0.100188	0.0489947	0.00203619	-0.0214541	-0.0669809	-0.142808	-0.169248	0.184151	1	0.145869
7	-0.00412765	-0.00640371	-0.00260146	0.0213644	0.0566404	0.0986235	0.163028	0.205627	0.0493263	0.0480445	0.0485058	0.0346981	0.0358952	0.0816254	0.0727846	0.0396299	0.00816948	-0.00530476	0.000815721	-0.010135	-0.0315226	-0.0742702	-0.122152	0.145869	1
	1	2	3	4	5	6	7	8	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	6	7

Table 1: Statistical correlation matrix between the absolute differential cross-sections. All variables are included to show the correlations between different bins of different variables. From left to right and top to bottom the rows and columns are labeled by bin number for each variable and the variables are ordered: p_T^t, $|y_t|$, $|y_{t\bar{t}}|$, and $m_{t\bar{t}}$.

The determination of statistical correlations within each spectrum and among different spectra are evaluated using the Bootstrap Method [22]. The method is based on the extraction of $\mathcal{N}$ Bootstrap samples from the data sample. The $i-th$ sample is made by associating a Poissonian weight to each event in data. From each Bootstrap sample the spectra are replicated following the very same procedure used for the nominal results. Since the weights are generated on an event-by-event basis, the replicated spectra are synchronized, thus allowing the determination of statistical correlations among different spectra.

The statistical correlations are evaluated bin-by-bin following the master formula:

$$C_{ij}^{AB} = \frac{\frac{1}{\mathcal{N}} \cdot \sum_{k=1}^{\mathcal{N}} (\mathcal{R}_i^{A,k} - \mu_i^A)(\mathcal{R}_j^{B,k} - \mu_j^B)}{\sigma_i^A \cdot \sigma_j^B}$$

where C_{ij}^{AB} is the element (i,j) of the statistical correlation matrix among spectra A and B, μ_i^A and σ_i^A are the mean and the standard deviation between the replicas in the i-th bin of spectrum A, respectively, and $\mathcal{R}_i^{A,k}$ is the content of the i-th bin of the k-th replica for spectrum A. The number of replicas has been set to $\mathcal{N}$ =100k.

Predictions for HERA DIS and ATLAS W,Z and Top

The formalism to relate PDFs to the DIS cross sections is text book stuff we only have to define the input PDFs and standard programmes do the rest

- QCDNUM for DGLAP evolution at NNLO
- DIS matrix-elements also from QCDNUM with RTVFN heavy quark scheme
- W,Z matrix elements at NLO from MCFM using Applgrid for input to PDF fit
- Augmented with NNLO/NLO k-factors from DYNNLO cross-checked with FEWZ for ATLAS, arXIV:1612.03016
- NLO-EW and photon induced corrections also applied

For top

Mitov et al issued fast grids at NNLO: arXiv:1704.08551 to facilitate PDF fitting using FastNLO. These can be used for the lepton+jets channel

For the dilepton channel MCFM NLO Applgrids are used with NNLO/NLO k-factors from arXiV:1611.08609

Mitov et all also issued Electroweak corrections arXIV: 1705.04105 these are included as k-factors

The predictions for y_t y_{ttbar}, m_{ttbar} are made for renormalisation and factorisation scale $H_T/4$, where

$$H_T = \sqrt{m_t^2 + (p_T^t)^2} + \sqrt{m_t^2 + (p_T^{tbar})^2}$$

Whereas the predictions for p_T^t use the scale $m_T/2$ where $m_T = \sqrt{m_t^2 + p_T^2}$

And m_t= 173.3 GeV.

These scale choices are taken from Czakon, Heymes, Mitov, arXIV:1606.03350

As usual in PDF fitting a parametrisation is assumed at a low scale Q^2_0

$$xq_i(x) = A_i x^{B_i}(1-x)^{C_i} P_i(x), \quad \text{where } P_i(x) = (1 + D_i x + E_i x^2) e^{F_i x}.$$

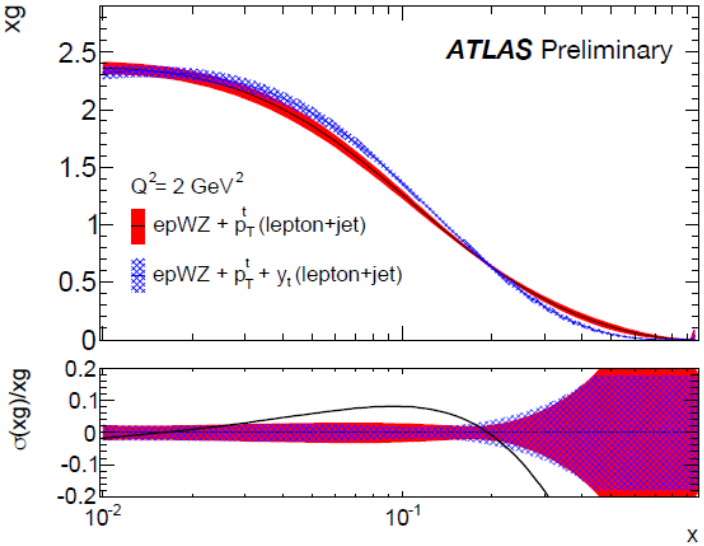
Where $xq_i(x)$ are the quark distributions $(xu_v,\ xd_v)$ and $(x\bar{u},\ x\bar{d},\ x\bar{s})$.

The gluon distribution has an extra term $xg(x) = A_g x^{B_g}(1-x)^{C_g} - A'_g x^{B'_g}(1-x)^{C'_g}$
Which allows larger uncertainties at small-x.
The valence and gluon normalisations are set by the number and momentum sum-rules
A few other constraints are applied to the low-x sea-such that ubar=dbar at very low-x,
But the strange normalisation is free --as for the ATLASepWZ16 fit.

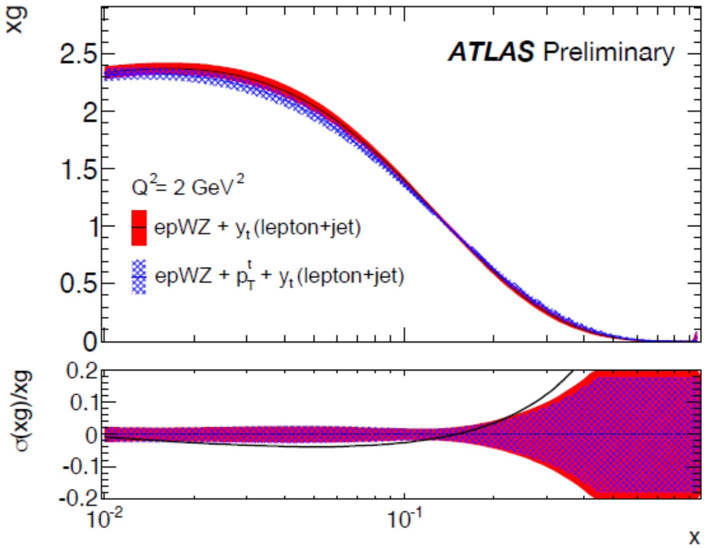
The fit begins assuming $P_i(x)=1$ and parameters D,E,F are added to each distribution until there is no further improvement in χ2---saturation of the χ2.
Some extra parameters can nevertheless change the shape of the PDFs and these are included as part of the parametrisation uncertainty.
Assumptions on the low-x sea are also varied as part of parametrisation uncertainty

PDF fits must also assume values for the starting scale Q^2_0=1.9GeV^2, the minimum Q^2=10GeV^2 of input data, the charm and beauty masses m_c=1.43, m_b=4.5 GeV and the strong coupling $\alpha_S(M_Z)$=0.118
All of the input values are varied as part of the model uncertainty

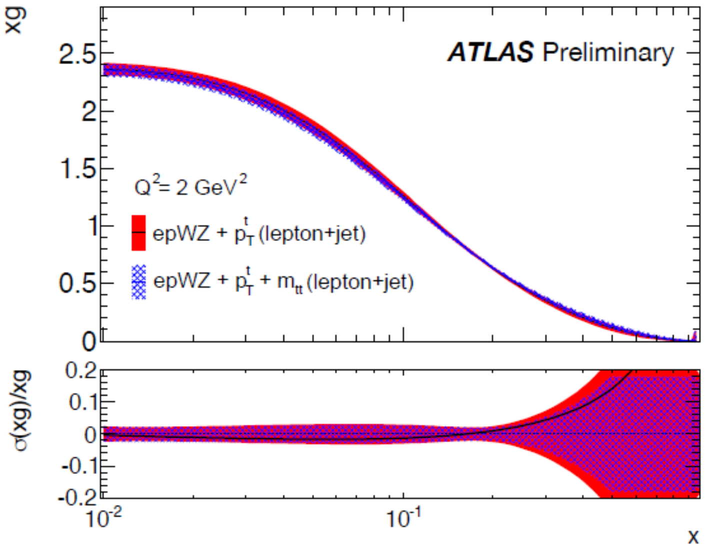
Now try the spectra two by two accounting for BOTH statistical and systematic correlations between the spectra



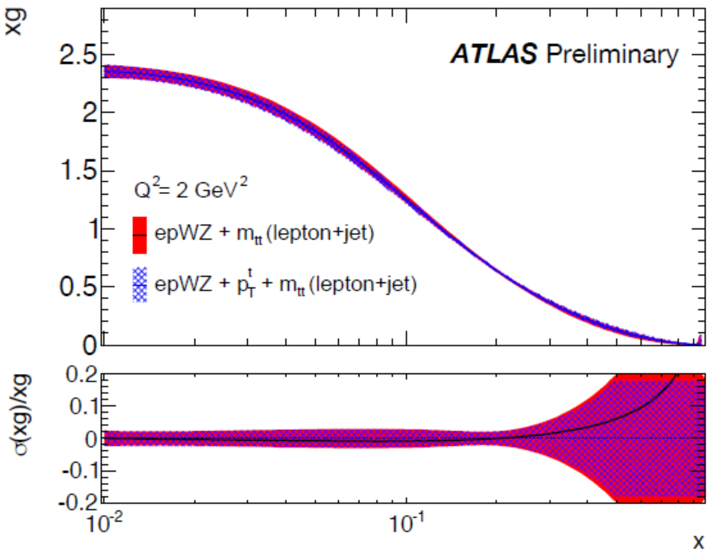
y_t has a stronger pull than $p_T{}^t$

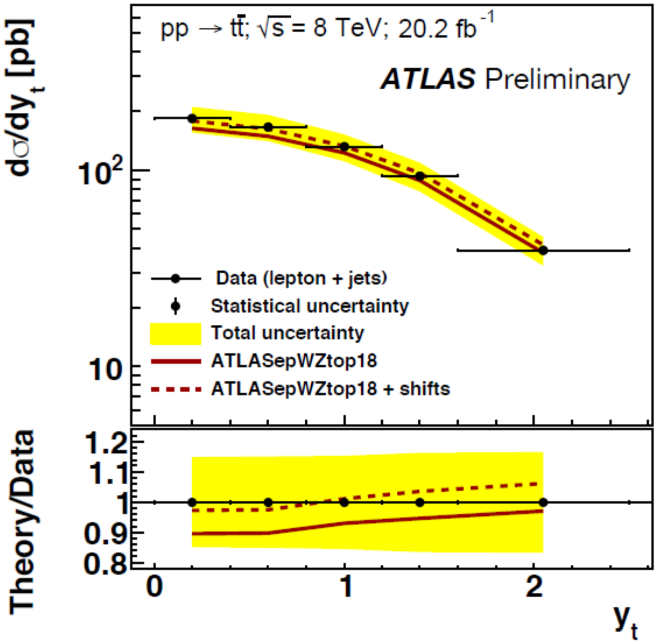


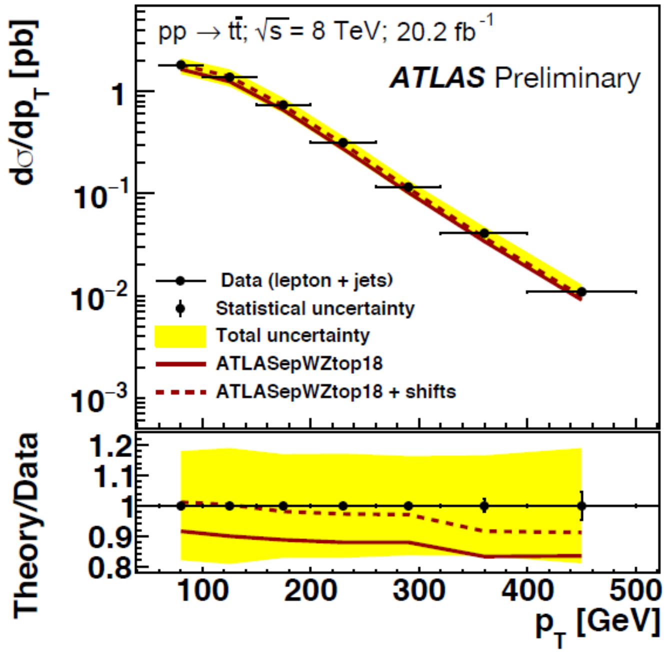
Compare fits first adding one spectrum and then two

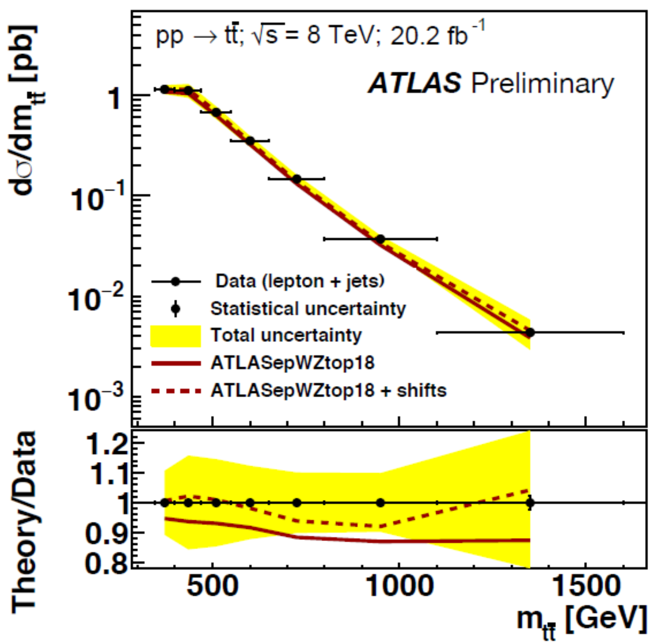


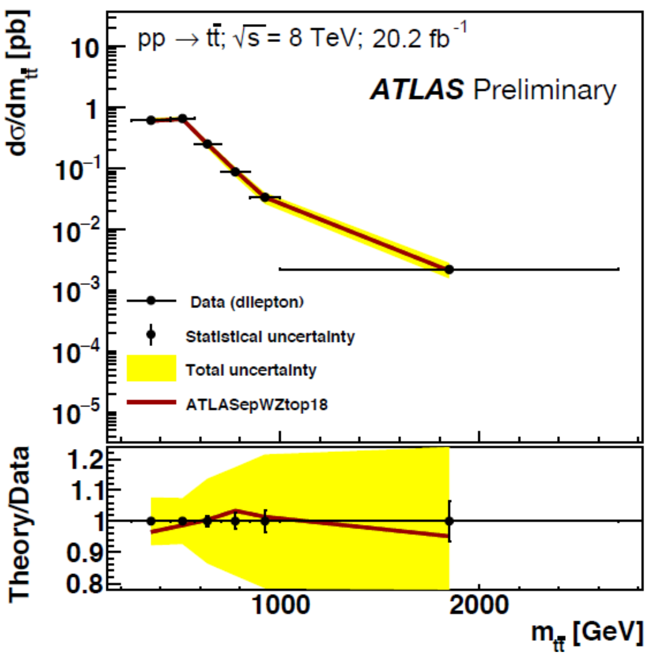
m_{ttbar} has a stronger pull than $p_T{}^t$











This one is not in the fit but is well described